DOCUMENT RESUME

ED 106 360 TM 004 476

AUTHOR Veldman, Donald J.; Sheffield, John R.
TITLE GUESS WHO: The Scaling of Sociometric Nominations.
PUB DATE Mar 75
NOTE 10p.; For related document, see TM 004 413; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Behavior; Comparative Analysis; Elementary Education; Elementary School Students; Emotional Development; Factor Structure; Group Relations; Individual Characteristics; Mental Retardation; *Neurologically Handicapped; Rating Scales; *Regular Class Placement; *Scoring Formulas; *Sociometric Techniques; Special Education; Test Reliability; Test Validity
IDENTIFIERS *Project PRIME

ABSTRACT

Two sociometric techniques were used in Project PRIME (Programmed Re-Entry Into Mainstream Education) to elicit data from peers about the behavior of selected normal and handicapped children in each of the 500 plus classrooms studied. One of these instruments was called Guess Who. The Guess Who instrument consists of 29 questions, such as, "Who is friendly to everyone?" Each child in the class writes the name of one classmate in response to each question. The names were encoded as identification numbers onto optical scanning sheets, which were then transfered to magnetic tape. Initial compilation of frequencies of nomination was carried out for 13,000 pupils. Four clear factors were obtained from analysis of 29 nomination sociometric items. Reliability estimates for the four factor scales are moderate, but adequate for comparison of groups. Educable Mentally Retarded (EMR) and Learning Disabled (LD) are perceived to have a definite cognitive deficit, and to be more disruptive in their behavior than normal peers. A pupil clustering technique yielded a stable seven group result. (Author/DEP)

ED106360

March, 1975

GUESS WHO: The Scaling of Sociometric Nominations*

Donald J. Veldman and John R. Sheffield

Project PRIME
Data Analysis Unit
The University of Texas at Austin

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Two sociometric techniques were used in Project PRIME to elicit data from peers about the behavior of selected normal and handicapped children in each of the 500 plus classrooms studied. One of these instruments was called Guess Who, and it is the reduction of these data that is the topic of the present paper.

The Guess Who instrument consists of 29 questions, such as, "Who is friendly to everyone?" Each child in the class writes the name of one classmate in response to each question.

The names were encoded as identification numbers onto optical scanning sheets, which were then transfered to magnetic tape.

Initial compilation of frequencies of nomination was carried out for 13,000 pupils. Classes of less than five pupils were dropped as instances of administration failure.

Factor Structure

Although the varying class sizes (from 5 to 46) do represent a serious methodological problem for the eventual computation of scale scores for pupils, there is no reason to expect class-size variation to distort the item correlation pattern, and hence the factor structure.

An image analysis of the 29 items resulted in a very clear four-factor structure.

---

* Presented at the 1975 meeting of the American Educational Research Association.

TM 004 476

Factor I was labelled "Disruptive." Two exemplary items are: "The teacher has to scold all the time," and "Is always bothering other children." Factor II was labelled "Bright." Two key items here were "Is the smartest in the class," and "Always knows the answers." Factor III was labelled "Dull." Two strong items were "Never knows the answers in class," and "Learns new things very slowly." The fourth factor was labelled "Quiet." Two important items were "Does not talk much to other children," and "Is the best behaved."

As a check on the assumption that class size does not influence the factor structure, data from the 1700 pupils in classes smaller than 16 were separately factored. The structure was virtually identical.

The content of these four factors suggests two fundamental bipolar dimensions: Academic Achievement and Misbehavior. The fact that each of these dimensions emerges as a pair of factors in the Guess Who analysis is a function of the nature of the instrument, which we will discuss later in this presentation.

Internal Consistency Reliability

Alpha coefficients were computed from successively longer Likert scales defined by the size of the factor loadings. The data indicated that the five highest-loading items on each factor constitute the most reliable scales for reduction of the instrument. For the four factors, these alpha coefficients were: .93, .93, .88, and .83.

We will return to the question of reliability later in this paper. For the present we will note only that the assumption of normal item distributions is seriously broken by the raw frequency data used to compute the alpha coefficients just reported.

The Class Size Problem

To enable interpretable statistical analysis of Guess Who scale scores, the meaning of a given score value must be comparable across classes of varying size. For low scores this is not a problem; it means the same thing not to be nominated at all in any size class. But a score of 15 in a class of 15 means 100% agreement; in a class of 30 the same score means only 50% agreement. High scores are obviously interpretable only with knowledge of class size.

Five different methods of scoring the Guess Who data were developed in an attempt to empirically determine an optimum procedure. These methods will be described, and then their relative validity against external criteria will be reported.

Raw frequency of nomination. This method was used in the factor analysis and alpha reliability analysis reported earlier. No attempt is made to adjust for class size. Correlations of these scores with class size are very low (.01, .03, .02, .04) simply because of the preponderance of low scores. It is the high scores that are of interest, however, and they are obviously strongly related to class size.

Proportion scores. It might seem that conversion of raw frequencies to proportions (of the class N) would neatly solve the problem. All this does, however, is shift the distortion from the high to the low scores. One nomination now becomes a score of 0.067 in a class of 15 and 0.033 in a class of 30. Because there are relatively so many low scores, the correlations of scale scores based on item proportions with class size are markedly negative (-.20, -.16, -.23, -.24). This method does not seem to hold much promise either.

Fixed-size panel. Suppose we had used a certain number of pupils (e.g., 15) in each class to do the nominating. This would seem to remove the class-size problem, since the maximum score in every class is the same. Such data were simulated by forming a fraction (15/N) to be used as a multiplier of each raw frequency score, then rounding the results to whole numbers. This method is obviously almost the same as proportion scoring. It also has the logical weakness of the fixed panel having more nominees available in larger than in smaller classes--leading to a greater expectation of low scores in larger classes.

Binary truncation. The raw frequency scores are converted to binary form (1=nominated one or more times, 0=not nominated). Because roughly 40% of the item scores are zeros, this conversion yields item-score distributions as close to normal as is possible. The apparent weakness of this method is the substantial loss of possibly useful information in the larger scores.

Standardization within classes. This is a straightforward way of equating the average scores of students in classes of various size. It has the drawback of moving a step away from the raw data, potentially introducing sample-specific error.

The first external validity analysis was a series of three-group, one-way analyses of variance to compare the diagnostic groups: 375 Educable Mentally Retarded, 205 Learning Disabled, and 1008 Normal Contrast children. The F-ratios for these comparisons showed that all methods were about equal in external validity, with the exception of the binary truncated scores, which produced F-ratios almost twice as large as those of any other method.

Also of interest was that the diagnostic groups were distinguished more clearly by the cognitive than by the behavioral scales. The negative cognitive scale also seemed to be more salient than the positive one.

The next series of analyses against external criteria were correlations of the five scoring systems against eight variables from two other sources. Four factor scales from a self-report instrument named "About You and Your Friends" were used, along with four factor scales from the "Teacher Rating Scales" instrument. The results of these analyses are as unequivocal as those of the analysis of variance. Correlations with the binary-truncated system scores were all substantially larger than with any other scoring system, when the relationships are clearly non-zero.

Generally, the correlations of comparable scales were much stronger with the teacher ratings than with the self-reports. For example, Guess Who "Disruptive" correlated .59 with teacher-rated misbehavior, but only .27 with self-rated misbehavior. Guess Who "Bright" correlated .51 with teacher-ratings of academic concentration, but only .23 with self-ratings.

Reliability Analyses

The alpha coefficients reported earlier were based on the use of raw frequency data, and the comment was made that the item distributions were badly skewed. The coefficients were recomputed using binary-truncated item data. For the four factors they were: .82, .77, .70, and .61.

It is now apparent that the reliabilities obtained with the raw frequency data were severely inflated due to failure to meet the assumption of normally distributed item scores.

Another approach to estimation of the reliability of the Guess Who scales is to split each of the classes of pupils into two panels of equal size and to compute scores for all pupils separately from item data in each nominating panel. This was done for 11,000 pupils in 400 classes with Ns between 16 and 37. The "split-class" reliabilities (not corrected with the Spearman-Brown formula) for the four factor scales were: .74, .72, .67, and .56.

Distribution of Classes by Size

The adjustment of scores for class size is one problem, but the meaningfulness of data from very small and very large classes is another.

Below a class size of 16, most of the pupils are classified MR or LD, which suggests that these are mostly self-contained special education classes, or regular classes in which the test administration process failed. The few very large classes are probably merged, team-taught arrangements.

It is doubtful that nominations by handicapped peers in a special class are comparable to nominations by mostly normal peers in a regular class. The meaningfulness of nominations in a very large merged class might also be questioned. For these reasons, subsequent analyses employed only data derived from pupils in classes of 16 to 37 pupils.

Scale Intercorrelation

The four factor scales, as noted earlier in this paper, seem to represent the ends of the two major bipolar dimensions similar to those measured by the Teacher Rating Scale: Academic Concentration and Misbehavior. The emergence of four unipolar factors from the Guess Who data, rather than two bipolar factors, may be considered an artifact of the nomination technique.

The essential feature of these scales is that a low score on a given trait does not imply a high score on its logical opposite. Lack of nomination as "bright" does not imply "dull." This peculiarity leads to expectations of stronger relationships across logical factors than within them. In fact, when the four scales were intercorrelated, this phenomenon emerged quite clearly. For instance, "Disruptive" correlated .44 with "Dull," but only -.26 with "Quiet."

Comparison of Diagnostic Groups

Let us return now to the analyses of variance comparing diagnostic groups, which was mentioned earlier. All four probability values were less than .0002. Omega-square values indicated that diagnosis is much more related to the cognitive dimension (14% of the variance) than to the behavioral dimension (2% of the variance).

The MR and LD children are very similar with regard to perceived cognitive deficit. The LD children are perceived as somewhat more disruptive than the MR children, however. The MR-LD difference is 50% of that between MR and NC children.

The NC children are nominated on the average about as much for "bright" as for "dull." The rate for handicapped children is about 5 to 1, however.

Typal Analysis

In any factor analysis there is always the possibility that some or all of the factors are better interpreted as types of people, rather than as traits possessed by all people in varying degrees. The four Guess Who factors could be construed as four types of children, although only the bright-dull and disruptive-quiet pairs appear to be mutually exclusive.

The question of types can be approached empirically by use of cluster analysis of pupil score profiles. The technique chosen is called hierarchical grouping analysis. Euclidian distances among standard-score profiles are used to build groups with maximum homogeneity. The step-wise group combination process continues until only two groups remain.

This procedure was carried out separately with the odd and even-numbered select subjects (N = 800 in each sample). The increases in within-group variance were inspected and the seven-group stage was chosen for further study.

The mean (Z-score) profiles of the seven groups in the pair of samples were highly similar. On the basis of their profiles the six types can be labelled (1) Disruptive, (2) Bright, (3) Dull, (4) Quiet, (5) Disruptive and Dull, (6) Bright and Quiet, and finally (7) Ignored.

There is a rather pleasing balance to this grouping. A type for each factor emerges, as well as the two combination types suggested by the inter-correlation pattern, and a final group of pupils who are essentially ignored--not nominated for any of the items consistently.

The obvious question to be asked next is with regard to the differential representation of the three diagnostic classifications among these seven sociometric syndromes. The seven-by-three frequency table yielded a highly significant chi-square value.

Perhaps the most surprising result was the large frequency of unnominated normal contrast children. Obviously, lack of nomination cannot be used as an operational definition of isolation.

The relative percentages of NC vs. MR-LD children in the other type groups were in line with intuitive expectations. Comparison of MR and LD composition of the three "negative" type groups shows them to be about equal for Dull, more likely MR for Disruptive, and more likely LD for Disruptive-Dull. This 2 by 3 comparison was not statistically significant, however.

Summary

In summary, we have reported that four clear factors were obtained from analysis of 29 nomination sociometric items. We have demonstrated that various methods of correcting the raw frequency data for class size are markedly inferior to simply converting each item score to binary form: nominated or not. We have noted that logically opposite ends of two dimensions appear as

four separate factors because of the nature of the nominations technique. As we have found in many other analyses of data from pupils and teachers, the two major dimensions appear to be cognitive performance and disruptive behavior.

Reliability estimates for the four factor scales are moderate, but adequate for comparison of groups. EMR and LD children are perceived to have a definite cognitive deficit, and to be more disruptive in their behavior than their normal peers. External validity of the scales against self-ratings is weak, but against teacher ratings it is substantial.

A pupil clustering technique yielded a stable seven-group result. Four groups were defined by single factors, two by pairs of factors, and the final group was composed of pupils who were not nominated for anything.